Combining Synthetic Controls and VARs:

On the Estimation of Causal Effects in Time Series Data

Jörg Breitung[a], **Lennart Bolwin**[b], **Justus Töns**[b]

[a]University of Cologne, Chair of Statistics and Econometrics
[b]University of Cologne, Chair of Statistics and Econometrics
Supervised by Prof. Jörg Breitung

Abstract

We argue that applications of Synthetic Control (SC) are faced with a self-selection problem. That is, the method is primarily applied to non-complex data structures that are straightforward to forecast, given the availability of donors in the post-treatment period. Using simulation studies, we show that the high interpretability of SC comes at the cost of poor predictions and forecasts, which are especially pronounced if the data generating process contains a time series structure. To address this issue, we introduce the intricacy-statistics that informs the applied researcher whether or not the data at hand exceeds a level of time series structure that SC can handle. If the case, more flexible methodologies that combine the strengths of SC and conventional time series techniques promise more accurate predictions and forecasts. Hence we introduce the new Vector Autoregressive Synthetic Control (VARSC) estimator, that takes in account both the time series structure and the availability of donors. In order to implement these ideas, we introduce the R-package varsc that provides ready-to-use functions to compute the intricacy-statistics and, based on the magnitude of the statistics, the functionalities to estimate either the SC or the VARSC model. To probe the performance of our methodology outside the experimental setting, we apply it to three existing applications of SC: Specifically, we show that our proposed model performs equally well like the SC-method. The result is striking because in contrast to the SC-model, our models gets along without the informational contend of potential covariates.

Keywords: *Synthetic Control; Causality; VAR*

List of Acronyms

ADH Abadie, Diamond, and Hainmueller

GDP Gross Domestic Product

DGP Data Generating Process

iid independent and identically distributed

MSFE Mean Squared Forecast Error

MSPE Mean Squared Prediction Error

OLS Ordinary Least Squares

PC Principal Components

SC Synthetic Control

USA United States of America

VAR Vector Autoregression

VARSC Vector Autoregressive Synthetic Control

Contents

List of Figures

1. Introduction

This is work in progress

SC method is combined with Vector Autoregression (VAR). Method was introduced by Abadie, Diamond, and Hainmueller (ADH). Paper strcuture could be similar to the one by [Doudchenko and Imbens, 2016]

2. Literature Review 2-3 pages

Literature has been clustered. This is work in progress

2.1. Synthetic Control

The SC method was developed by Alberto Abadie and colleagues in a series of influential papers ([Abadie and Gardeazabal, 2003], [Abadie et al., 2010], [Abadie et al., 2015]). The method is designed to estimate the causal effect of a treatment in a setting with a single treatment unit and a number of potential control units. Pre- and post-treatment data are observed for the treatment and control units for the outcome of interest as well as for a set of covariates. The SC-procedure combines aspects of the matching and difference-in-difference literature and can therefore be interpreted as a relative of the causal inference literature introduced by [Rubin, 1974]. Similar to many other microeconometric methods, the objective is to distinguish causation from correlation and to assess the magnitude and significance of treatments in observational case studies.

In their canonical 2003 article, Abadie and Gardeazabal evaluate the causal economic effects of conflict using terrorist conflicts in the Basque Country as a comparative case study. In their specific application example, they find that terrorist conflicts caused the per capita Gross Domestic Product (GDP) of the treatment unit (Basque Country) to decline by about 10% relative to the synthesized control unit.

some more words on other findis and things they did The next appropriate setting for an application of the SC method was the introduction of a large-scale tobacco control program implemented in the state of California in the United States of America (USA) in 1988.

2.2. Overview

[Abadie, 2021] read.

[Athey and Imbens, 2016] read.

2.3. Application

[Born et al., 2019] read.

[Cho, 2020] read.

[Cunningham, 2021] read.

[Funke et al., 2020] read.

2.4. Methodological Background

[Hainmueller et al., 2011] read.

[Abadie and Imbens, 2006] not read.

[Abadie and Imbens, 2002] not read.

[Doudchenko and Imbens, 2016] read.

[Ferman, 2021] read.

[Frangakis and Rubin, 2002] not read.

[Rosenbaum and Rubin, 1983] not read

[Rubin, 1974] not read.

2.5. Extensions/ Developments

[Abadie and L'Hour, 2021] read.

[Amjad et al., 2018] read.

[Ben-Michael et al., 2021] read.

[Ben-Michael et al., 2021] not read.

[Kellogg et al., 2021] not read.

[Kuosmanen et al., 2021] not read.

[Muhlbach and Nielsen, 2019] read.

Developments

[Arkhangelsky et al., 2021] not read

[Athey et al., 2017] not read.

[Brodersen et al., 2015] read.

[von Brzeski et al., 2015] read.

[Hartford et al., 2017] read.

2.6. Testing

[Andrews, 2003] not read.

[Cattaneo et al., 2021] not read.

[Chernozhukov et al., 2019] not read.

[Chernozhukov et al., 2021] not read.

[Firpo and Possebom, 2018] not read.

[Hahn and Shi, 2017] read.

2.7. Time Series Econometrics

[Martin et al., 2012] read.

[Harvey and Thiele, 2020] read.

[Breitung and Knüppel, 2021] partially read.

3. Theory

This is work in progress. Align with previous chapters when they are written

In this chapter, we propose an alternative SC-estimator to assess the magnitude of treatment effects in observational settings. To establish a general basis, let us first describe the contextual environment of the estimation. Similar to the setting as introduced by ADH, we consider a framework with $J+1$ panel units indexed by $j = 0, 1, ..., J$ that are observed over a time horizon of T periods. Without loss of generality, assume that unit $j = 0$, is exposed to the treatment at period $t = T_0$ with $1 < T_0 < T$ and that there are no treatment anticipation and contamination (i.e., no spillovers in time and space). The former would be the case if the treatment affects unit $j = 0$ before T_0, the latter describes the case where some of the supposedly untreated units $j = 1, ..., T$ are contaminated as they are affected by the treatment. To contextualize these assumptions, [Abadie et al., 2010] argue that in the presence of anticipation effects, T_0 could be shifted back in time until the assumptions seems plausible. If panel units in the donor pool[1] are affected by the treatment (contamination) as it is likely in Brexit-application, those units could be removed from the sample prior to the estimation. Our goal is to evaluate the causal effect of the treatment, the specific functional form of which remains unspecified though. This is possible because the main goal of the SC-estimation lies in the precise estimation of the counterfactual. Since the treatment scenario is empirically observable, it is not necessary to specify the functional form of it (e.g. level or slope shift, fading or persistent shock).

The following theoretical argumentation is structured as follows: We first describe the canonical estimation procedure as proposed by ADH. Next we build intuition by considering a simple static scenario with only two donor units and one treatment unit that is then generalized to the case with many donors. The main difference between our extensions and the setting of ADH is that we remove some of the weight constraints and that we analyze a situation without covariates. The former distinction guides us to the field of regularization in order to prevent our method from overfitting. The latter drastically

[1] To ensure direct comparability with the SC literature, we adopt most of the commonly used terms. For example, control group units are labeled as 'donors'.

reduces the data requirements but causes our algorithm to estimate the counterfactual with a significantly smaller information set. This fact leads us to our main contribution: The integration of multivariate time series approaches into the SC-algorithm.

3.1. ADH Case

We start by presenting the SC-method in its original form as introduced ADH. Besides introducing the general estimation technique, we also want to elaborate on the proposed hypothesis testing procedure of ADH. For the sake of comparability and due to its notational and inhat clarity, we borrow the employed notation of Abadie and colleagues. In terms of the structural design, we build on the thorough presentation of SC and the proposed hypothesis testing procedure by [Firpo and Possebom, 2018].

Setup

The estimation task can be constituted by the potential outcome framework as introduced by [Neyman, 1923] and elaborated by [Rubin, 1974]. Let $Y_{j,t}^I$ be the (potential) outcome for unit j at point t in the presence of the intervention. Likewise, let $Y_{j,t}^N$ be the (potential) outcome for j at point t in the absence of the intervention. ADH define the treatment effect of the intervention as

$$\delta_{j,t} = Y_{j,t}^I - Y_{j,t}^N$$

and introduce the indicator variable $D_{j,t}$ that takes on the value 1 if unit j is treated at period t and the value 0 otherwise. Given the assumed absence of anticipation and contamination, we observe the following outcome

$$Y_{j,t} + D_{j,t}\delta_{j,t} = \begin{cases} Y_{j,t}^N & (\text{if } j = 0 \text{ and } t < T_0) \text{ or } j \geq 1, \\ Y_{j,t}^N + \delta_{j,t} & \text{if } j = 0 \text{ and } t \geq T_0 \end{cases}$$

The goal to estimate the causal treatment effect $(\delta_{0,T_0}, ..., \delta_{0,T})$ therefore boils down to the estimation of the counterfactuals of unit $j = 0$ in the post-treatment phase $(Y_{0,T_0}, ..., Y_{0,T})$, i.e. on what trajectory would unit $j = 0$ have beeen had there been no intervention. The basic idea of ADH is to estimate these counterfactuals as a weighted average of the donor outcomes, using a data-driven approach to compute the weights. Intuitively, the weights

are computed such that they optimally predict the outcomes and a set of explanatory variables of the treatment unit in the pre-intervention phase, conditional on having a percentage interpretation. To operationalize this intuition, let $\boldsymbol{Y_j} = (Y_{j,1}, ..., Y_{j,T_0})'$ be the vector of observed outcomes in the pre-treatment phase for unit j[2]. To distinguish treatment unit and donors, ADH denote the $(T_0 \times 1)$-vector of the treatment unit as $\boldsymbol{Y_1}$ and the $(T_0 \times J)$-matrix of the donors as $\boldsymbol{Y_0}$. Moreover, a set of K covariates is observed for all panel units for $t = 1, 2, ..., T$, yet only the pre-treatment values are needed for the weight-calculation. Therefore, let $\boldsymbol{X_1}$ denote the $(K \times 1)$-vector of covariates for $\boldsymbol{Y_1}$ and let $\boldsymbol{X_0}$ denote the $(K \times J)$-matrix of explanatory variables for $\boldsymbol{Y_0}$. In order to estimate the causal effect of the treatment, the SC-estimator estimates the counterfactuals $(\widehat{Y}_{0,1}, ..., \widehat{Y}_{0,T_0}, ..., \widehat{Y}_{0,T})$ for the pre- and post-intervation phase as

$$\widehat{\boldsymbol{Y}}^N_{0,t} = \sum_{j=1}^{J} \widehat{w}_j Y^N_{j,t} \; \forall \; t \in \{1, ..., T\}$$

The weights are stacked in the vector $\widehat{\boldsymbol{W}} = (\widehat{w}_1, ..., \widehat{w}_J)'$ and have the constraint to have a percentage interpretation such that $\widehat{w_j} \geq 0$ and $\sum_{j=1}^{J} \widehat{w_j} = 1$. It is worth noting that the percentage interpretation of the weights requires the counterfactuals to belong to the convex hull of the donors as otherwise, $\widehat{\boldsymbol{Y}}^N_{0,t}$ will never match its true counterpart. [Abadie et al., 2010] argue that "the magnitude of discrepancy" should be calculated in advance of each SC-application. If the researcher finds that the pre-intervention values of $\widehat{\boldsymbol{Y}}^N_{0,t}$ fall outside the convex hull of the donors, the employment of SC is not recommended. The weights $\widehat{\boldsymbol{W}}$ are obtained as the solution of a nested optimization problem that aims to match both the pre-treatment outcomes $(\boldsymbol{Y_1})$ and a set of fixed pre-treatment covariates for the treatment unit $(\boldsymbol{X_1})$. ADH formalize this idea as follows

$$\widehat{\boldsymbol{W}}(\boldsymbol{V}) = \underset{\text{s. t. } \widehat{w_j} \geq 0 \text{ and } \sum_{j=1}^{J} \widehat{w_j} = 1}{\arg\min} (\boldsymbol{X_1} - \boldsymbol{X_0 W})' \boldsymbol{V} (\boldsymbol{X_1} - \boldsymbol{X_0 W})$$

with $\boldsymbol{V}$ being an arbitrary diagonal positive semidefinite weight matrix of dimension $(k \times$

[2] For instance, in the canonical example of [Abadie and Gardeazabal, 2003], $\boldsymbol{Y_j}$ would be the vector of GDPs for panel unit j.

k). $\boldsymbol{V}$ itself is the solution of the following optimization problem

$$\widehat{\boldsymbol{V}} = \underset{\text{s. t. } \widehat{\boldsymbol{V}} \in \mathcal{V}}{\arg\min} (\boldsymbol{Y_1} - \boldsymbol{Y_0}\widehat{\boldsymbol{W}}(\boldsymbol{V}))'(\boldsymbol{Y_1} - \boldsymbol{Y_0}\widehat{\boldsymbol{W}}(\boldsymbol{V}))$$

with $\mathcal{V}$ being the set of all positive semidefinite weight matrices of dimension $(k \times k)$. Subsequently, the causal effect of the intervention $\delta_{j,t}$ can be quantified at each time point after the intervention $t = t_0+1, ..., T$ as the gap between observed $(Y_{0,t}^N+\delta_{j,t})$ and predicted outcome $(\widehat{Y}_{0,t}^N)$.

This two-step estimation procedure serves two crucial purposes: $\widehat{\boldsymbol{V}}$ measures the relative importance of the variables in $\boldsymbol{X_0}$ to explain $\boldsymbol{X_1}$. Specifically, $\widehat{\boldsymbol{V}}$ is selected such that it minimizes the Euclidian distance of the pre-intervention outcome of unit $j = 0$ and its synthesized counterpart defined by $\widehat{\boldsymbol{W}}(\boldsymbol{V})$. The weighting vector $\widehat{\boldsymbol{W}}(\boldsymbol{V})$ in contrast quantifies the relative importance of each unit in the donor pool. Summarizing the key concept of ADH, the SC-method ensures that the synthesized treatment unit is as similar as possible to the actual treatment unit with respect to the quantity of interest and a set of potential explanatory variables in the pre-treatment period. Especially in the canonical examples of SC, the quantity of interest (e.g. GDP) and the explanatory variables (e.g. investment, savings etc.) were inherently interconnected. Thus, observing that the SC-estimator was capable of approximating both targets significantly enhanced the methods credibility. If explanatory variables are omitted, the SC-algorithm reduces to an Ordinary Least Squares (OLS) estimation, constraint such that the constant equals zero and that the coefficients have a percentage interpretation.

The main concern when assessing treatment effects with the SC-method is the poor generalizability of the estimation results in the post-treatment period. For example, especially when employing non-parametric statistical learning methods, it is simple to achieve high in-sample (pre-treatment) fit. The crucial part when dealing with forecasts is that the observed in-sample patterns generalize well outside the verifiable horizon (post-treatment). One way to assess methods for detecting generalizable patterns is through hypothesis testing.

Hypothesis Testing

ADH propose a model-invariant non-parametric inference procedure that is based on the Exact Hypothesis Test proposed by [Fisher, 1935]. The basic idea behind such permutation tests is to compare the observed data with a number of randomly permuted versions of the data, and to use the distribution of the test statistic calculated from these permuted samples to estimate the probability that the observed result occurred by chance alone.

In the context of SC, ADH consider permutations in region (i.e. panel unit) and time. Region permutations estimate the treatment vector $(\delta_{j,T_0}, ..., \delta_{j,T})$ for each panel unit $j \in \{0, ..., J\}$.[3] This procedure provides them with the empirical $(J+1)$-observational distribution of the treatment. Subsequently, it is possible to compare the estimated treatment vector $(\delta_{0,T_0}, ..., \delta_{0,T})$ of the treated unit with the J placebo-treatment vectors of the units of the donor pool. Given the estimated treatment effect for $j = 0$ is large, the null hypothesis of no treatment effect can be rejected at the significance level of one minus the percentile of $(\delta_{0,T_0}, ..., \delta_{0,T})$ in the empirical distribution.[4] Time permutations on the other hand consider only panel unit $j = 0$, permute T_0 to dates prior to the true treatment date and again compute the empirical treatment distribution. Given that $T_0 > J$, this approach can increase the sensitivity of the test, since the theoretically feasible significance threshold of region permutation tests is determined by $\frac{1}{J}$. For both, region and time permutations, ADH condense the vector of estimated treatment effects into a precision metric like the Mean Squared Forecast Error (MSFE)[5] of the following form:

$$MSFE_j = \frac{\sum_{t=T_0}^{T} \left(\widehat{Y_{j,t}^N} - Y_{j,t}^N\right)^2}{T - T_0}$$

A potential problem that may arise when examining the relative rarity of the estimated treatment effect using the procedure described above is inherent in the nature of SC. In the context of region permutations, suppose that a donor region is very different from the rest such that it falls outside the convex hull of the remaining donors. Note, that

[3] Note that it is necessary to exclude the truly treated unit from donor pool to ensure the validity of the no contamination assumption.

[4] For instance, let $J = 99$ such that treatment effects for 100 panel units can be computed. As long as the estimated treatment effect of the truly treated units belongs to the 95 largest effects (95th percentile or higher), the permutation test rejects the null hypothesis of no treatment effect at least at 5 percent.

[5] Note that ADH speak of the Mean Squared Prediction Error for dates before and after T_0. Since we consider the time span until T_0 as prediction window and the time span after T_0 as forecast window, we employ the label Mean Squared Prediction Error (MSPE) before T_0 and the label MSFE from T_0 onward.

this circumstance does not cause problems for the truly treated region and its synthesized counterfactual. The outlier described above is likely to be assigned a weight of zero in the estimation of the potential outcome without treatment. However, since the outlier cannot be synthesized precisely by the donor pool by construction, both MSPE and MSFE are expected to be large. As this special feature causes the permutation test to be unreasonably conservative, ADH propose to exclude regions who are hard to predict, i.e. who have a MSPE that exceed the MSPE of the truly treated unit to a great extent.

Figure 1 visualizes the exclusion procedure in the tobacco control application of ADH. The vertical axis indicates the gap between observed and estimated per capita cigarette sales, with the bold line representing the truly treated region (California). Looking at panel A, two observations stand out: First of all, the treatment has a clear negative effect for California. Second, some regions have both a poor pre- and post-treatment fit. Since the estimated treatment should not be artificially driven by a poor fit, ADH successively remove regions with a large MSPE relative to California. Panel B excludes regions with a MSPE that is more than 20 times as large the MSPE of California, Panel C lowers the cutoff to five times California's MSPE and Panel D to two times the MSPE. In the last scenario, only 19 regions are left and California is the one with the most extreme treatment effect. The treatment is therefore statistically significant with a p-value of 5,3% $\left(\frac{1}{19}\right)$.

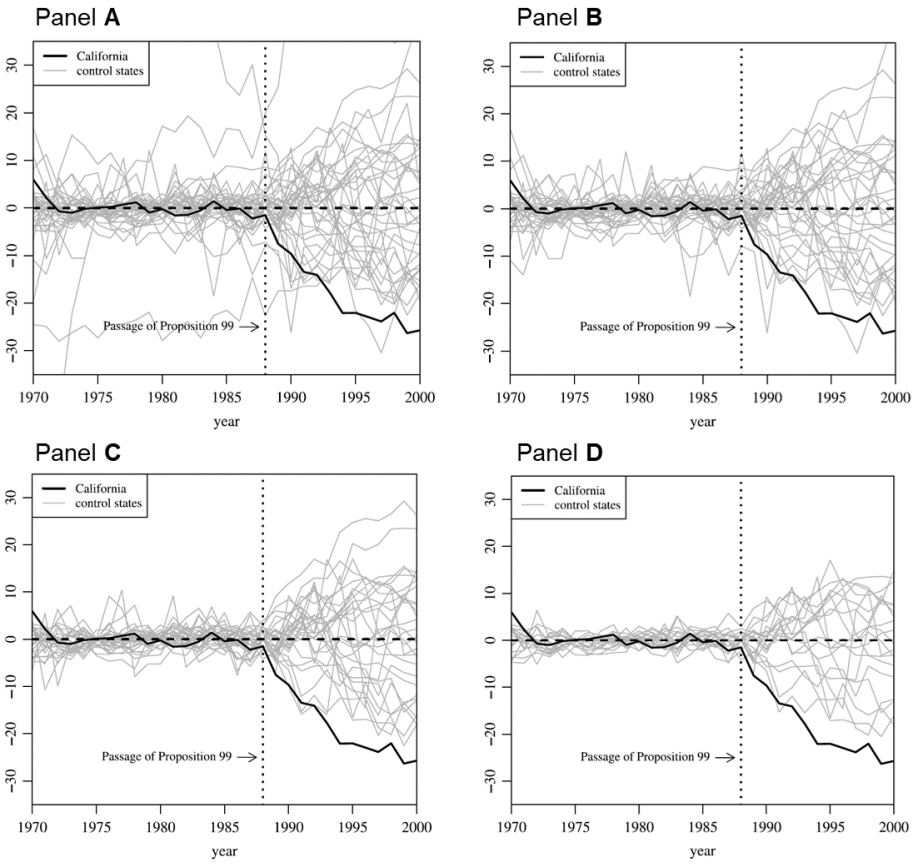

Figure 1. Region Exclusion Procedure of ADH

One way to bypass the inefficient sample reduction procedure is to look at the distribution of the ratios of MSFE and MSPE. By scaling the post-treatment fit with the pre-treatment fit, regions with a poor fit are implicitly controlled for. In the tobacco control application, California is the region with the highest MSFE-to-MSPE ratio among all 39 regions which translates into a p-value of 2,6% $\left(\frac{1}{39}\right)$.

3.2. Simple Static Extension

This is work in progress

to dos

- **consistent notation**

Consider a very simple framework for analyzing the causal effect of a treatment for unit $i = 0$ and two units in the control group $i = 1, 2$. It is assumed that before the

intervention at time period $t = T_0$ the units have a joint distribution of the form

$$\boldsymbol{y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ before } T_0.$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)'$ and $\boldsymbol{\Sigma}$ is some positive definite covariance matrix with Choleski decomposition $\boldsymbol{\Sigma} = \boldsymbol{R}\boldsymbol{R}'$ and $\boldsymbol{R}$ is an *upper* triangular matrix. Assume that the intervention affects the mean of the first variable such that $\mathbb{E}(Y_0) = \mu_0 + \delta$ after the intervention, whereas the means of the other two variables remain unaffected. Accordingly, δ represents the treatment effect on Y_0.

We are interested in deriving an optimal estimator for the counterfactual

$$\widehat{Y}_0^N = \mathbb{E}(Y_0 | \delta = 0, Y_1, Y_2) \text{ after } T_0.$$

Let $Q = R^{-1}$ and q denotes the first row of Q, then

$$q'\boldsymbol{y} = q'\boldsymbol{\mu} + \epsilon,$$

where $\epsilon \sim \mathcal{N}(0, 1)$ with $\mathbb{E}(\epsilon | Y_!, Y_2) = 0$. It follows that

$$\begin{aligned} \widehat{Y}_0^N &= w_1 Y_1 + w_2 + Y_2 + \mu^* \\ &= \mu_0 + w_1(Y_1 - \mu_1) + w_2(Y_2 - \mu_2), \end{aligned}$$

where $w_1 = -q_1/q_0$ and $w_2 = -q_2/q_0$ and $\mu^* = \mu_0 - w_1\mu_1 - w_2\mu_2$. These results imply that there is no reason to impose the restrictions $w_1 \leq 0, w_2 \leq 0$ (positivity) and $w_1 + w_2 = 1$ (adding-up). Furthermore, the construction of SC should include a constant term, as otherwise the SC may have a different mean, See also [Doudchenko and Imbens, 2016] for a careful discussion of these restrictions.

For illustration assume that

$$\boldsymbol{y} \sim \mathcal{N}\left(\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1 & 0.1 & 0.4\\0.1 & 1 & 0.5\\0.4 & 0.5 & 1\end{pmatrix}\right)$$

Elaborate here once understood. For this example the optimal weights for the SC result as $w_1 = -0.133$, $w_2 = 0.4667$ and $\mu^* = 1 - w_1 - w_2 = 0.667$. Note that w_1 is negative even all bivariate correlations between the panel units are positive. One may argue that this solution does not make much sense as from a economic perspective it is not clear what it means that Y_1 enters the SC with a negative sign. This demonstrates the trade-off between optimality in a statistical sense and the interpretability of the solution.

What happens if we impose the restrictions that all weights are positive and sum up to unity? In this case the restricted optimum yields the linear combination $\widetilde{Y}_0^N = 0.2Y_1 + 0.8Y_2$. The important difference lies in the variance of these estimates. For our example we obtain

$$var(Y_0 - \widehat{Y}_0^N) = 0.827$$
$$var(Y_0 - \widetilde{Y}_0^N) = 1.160$$

It is interesting to note that the variance of the restricted estimate is even larger than the unconditional variance of Y_0. This is possible as $(w_1, w_2) = (0, 0)$ is not included in the restricted parameter space.

It is not difficult to see that if Y_0 is not correlated with Y_1 and Y_2, then the optimal estimate boils down to $\widehat{Y}_0^N = \mu_0$ and therefore it does not make sense to involve a SC. In microeconometric settings it is usually assumed that the individuals in the treatment group and individuals in th econtrol group are uncorrelated. In such cases we do not care about constructing a SC. The crucial feature of SC methods is the correlation between the units in the treatment and the control group. In macroeconomic applications however, the variables in the treatment and control group (e.g. GDP) are typically correlated and it is therefore important to model the relationship between the variables. As the simple scenario with only two panel units in the donor pool is highly unrealistic in practice, we

now move to the general static case with $k-1$ panel units.

3.3. General Static Extension

What must be clear by now

- **Derive first analytical expressions for the case with k donors before talking about regularization**

In empirical practice it is often the case that the number of pre-intervention time periods $T_0 - 1$ is small and my even be smaller than the number of units in the donor pool, k.

3.4. General Dynamic Extension

What must be clear by now

- **TBD**

When modeling macroeconomic time series it is often assumed that the $(k+1) \times 1$ vector of time series $y_t = (Y_{0t}, ..., Y_{kt})'$ can be represented by a VAR model given by

4. Simulation

4.1. Static Data Generating Processes

Simulation-Procedure

- [Abadie et al., 2010] suppose that the counterfactuals $Y_{1,t}^N$ for $t > T_0$ are given by a factor model of the form $Y_{i,t}^N = \alpha_t + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{it}$. α_t denotes an unknown common factor with constant loadings across panel units. Z_i is a vector of observed panel-specific covariates, θ_t is a vector of unknown parameters, λ_t is a vector of unknown common facots and μ_i are panel-specific unknown factor loadings. The unobserved transitory shocks ϵ_{it} have zero mean at the panel level. For this specific setting, [Abadie et al., 2010] show that "[...] the bias of the SC-estimator can be bounded by a function that goes to zero as the number of pre-treatment periods increases." Further the number of donor units has to be fixed.

- [Ferman, 2021] considers a de-meaned scenario without additional covariates. The representation of the counterfactuals therefore boils down to $Y_{i,t}^N = \lambda_t \mu_i + \epsilon_{it}$. Thus the counterfactual is given by the composition of the unknown panel-specific factor loadings and the unknown common factors plus the idiosyncratic shocks.

- We re-estimated the Tobacco Control and the Basque application and found that the inclusion of additional covariates did not improve the predictive accuracy of the SC estimator. Therefore, we follow the simulation suggestion of [Ferman, 2021] and root our simulation in his proposed factor model without additional covariates.

- Analogous to Ferman, we consider a setting with two common factors, $\lambda_{1,t}$ and $\lambda_{2,t}$. The potential outcomes for the treated unit and for the first half of the donor pool load exclusively with loading one on the first factor, the remaining donors load exclusively with loading one on the second factor. Therefore μ_i is a (2×1)-rowvector with the first (second) entry being one and the second (first) entry being zero for the first (second) half of the donor pool.

- In each simulation, we simulate a total of $T_0 + T_1$ observations, where T_0 represents the length of the pre- and T_1 the length of the post-treatment period. In order to

have a simulation framework that is as close as possible to real world SC-applications, we choose $T_0 \in \{20, 50, 100\}$ and $T_1 \in \{10, 20, 30\}$. Further, we define the size of the donor pool as $J \in \{5, 10, 15, 20, 25, 30\}$.

- To initialize the simulation, we simulate the two common factors by drawing $T_0 + T_1$ independent and identically distributed (iid) observations from a standard normal distribution. Next, we multiply the common factors for all donors and the treatmend unit with the factor loadings μ_i and add iid standard normal shocks. Additionally, we add panel-specific iid standard normal intercepts such that our DGP takes the following form: $Y_{i,t}^N = \gamma_i + \lambda_t \mu_i + \epsilon_{it}$.

- Figure 2 visualizes one potential DGP with $T_0 = 20$ and $T_1 = 10$. To make the factor structure visible, we scaled the factor variance by 10^1 and the shock variance by 10^{-1}. Moreover, for better eyeball inspection, we added a constant treatment effect of $\delta_{1,t} = 10$ for $t > T_0$. Note that the specific nature of the treatment effect is irrelevant for our investigation. Since $Y_{1,t}^T$ is observable for $T > T_0$, we can choose an arbitrary functional form or even implement no treatment effect at all.

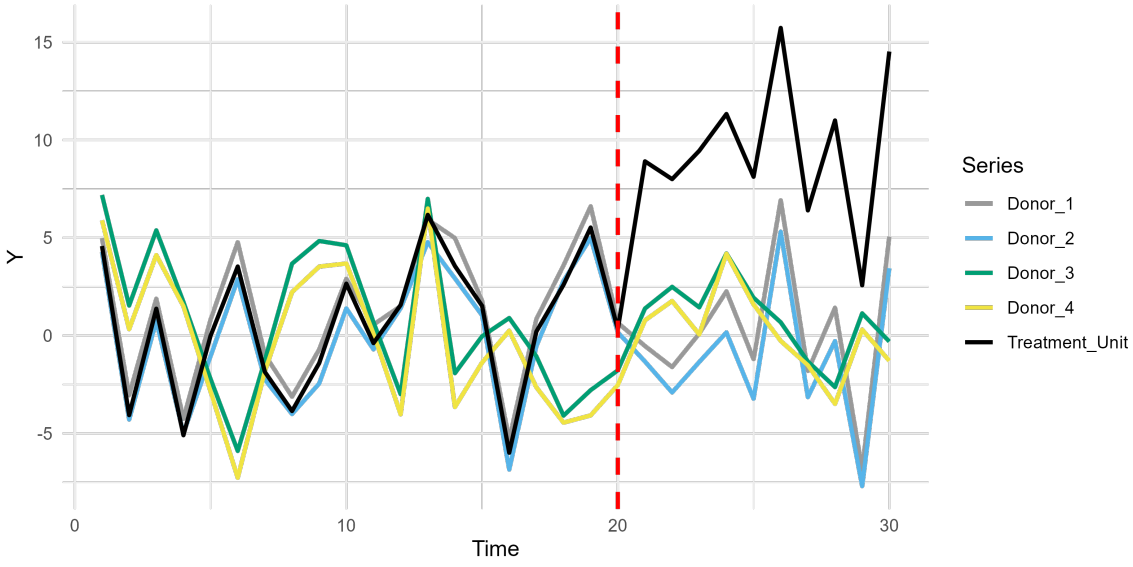

Figure 2. Example Factor-DGP

- The described factor structure is inherent in Figure 2: The treatment unit and the first half of the donors (Donor_1 and Donor_2) as well as the second half of the

donors (Donor_3 and Donor_4) share a common factor structure. The constant treatment effect is added after $T_0 = 20$ periods, resulting in a 10 unit vertical shift of the treatment series.

- We consider five different models whose goal is to recover the true factor structure of the treatment unit in the pre-treatment period and to predict the counterfactual in the post-treatment period as accurately as possible. Put differently, we train the models on T_0 pre-treatment observations and test their performance using T_1 post-treatment observations.

- The following models are employed: **(Update when previous parts are written)**

  1. Factor-Model (FACTOR): It is assumed that the common factors and the idiosyncratic component are uncorrelated and the idiosyncratic errors are mutually uncorrelated. This suggests to estimate the common factors as linear function of the donors. A popular estimator with this property is the Principal Components (PC) estimator. In the training period, we obtain the predictions by regressing the treatment series on the "latent" factors. As we implemented a two-factor structure, the factors are computed by multiplying the first two eigenvalues of the covariates covariance matrix with the matrix of the covariates. The forecasts for the testing period are obtained by multiplying the factor structure of the post-treatment period with the regression coefficients of the pre-treatment regression. This model is most natural when generating data according to a factor process. Therefore, we expect it to perform best among all models.

  2. Synthetic-Control-Model (SC): In case of no additional explanatory variables, the SC-approach reduces to a constraint regression that regresses the treatment series on the donor series given the constraint of no intercept and non-negative coefficients that sum up to one. ADH argue that the constraints prevent the SC models from over-fitting, but they do so at the cost of a reduction in flexibility.

  3. Regularized OLS-Model (REGOLS): Our proposed model. We allow for a constant and arbitrary coefficient values. We propose a two-dimensional regularization: A ridge/l2-norm penalty that penalizes the coefficient sum and a

second penalty term that shrinks the coefficient sum towards one. We identify the optimal hyperparameter-combination by applying a 50/50 train-test-split in the pre-treatment period on a two-step random grid search. In the first step, we randomly select 400 hyperparameter-combinations from a (50×50)-grid. In the second step, we enclose the potential optimum by sequentially holding the first and the second hyperparameter fixed while increasing and decreasing the remaining hyperparameter on a coarser grid.

4. Elastic Net (NET): [Doudchenko and Imbens, 2016] propose to estimate the counterfactual using an elastic net regression.

5. OLS-Model (OLS): Finally, to verify the belief that a simple linear model overfits the training data and therefore yields poor forecasting accuracy in the testing period, we also employ an ordinary least squares model.

All simulations are conducted with 1,000 iterations per combination of pre- and post-treatment horizon and donor quantity.

Table S1. Simulation Results of the Static Factor Model with **J = 5** Donors.

T_0	T_1	**FACTOR**	**SC**	**REGOLS**	**NET**	**OLS**
		RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]
20	30	1.2494 (0.0166) [0.6158]	1.4604 (0.0001) [1.1078]	1.2942 (0.0142) [0.6464]	1.2959 (0.0123) [0.5695]	1.3552 (0.0102) [1.1093]
	20	1.2595 (-0.0068) [0.6477]	1.4560 (0.0320) [1.0652]	1.3080 (-0.0009) [0.6610]	1.3067 (-0.0003) [0.5948]	1.3665 (0.0066) [1.0979]
	10	1.2342 (0.0250) [0.6312]	1.4322 (0.0256) [1.0473]	1.2877 (0.0256) [0.6361]	1.2865 (0.0202) [0.5516]	1.3569 (0.0309) [1.0745]
50	30	1.1819 (-0.0025) [0.6326]	1.4327 (-0.0015) [1.0579]	1.2130 (-0.0036) [0.6444]	1.1987 (-0.0014) [0.5430]	1.2133 (-0.0034) [0.7755]
	20	1.1837 (0.0092) [0.6469]	1.4224 (0.0097) [1.0563]	1.2124 (0.0144) [0.6396]	1.2046 (0.0131) [0.5583]	1.2194 (0.0150) [0.7840]
	10	1.1663 (0.0038) [0.5825]	1.3990 (-0.0019) [0.9754]	1.1966 (-0.0041) [0.5731]	1.1815 (0.0045) [0.5071]	1.1968 (-0.0014) [0.7163]
100	30	1.1636 (-0.0046) [0.6505]	1.3944 (0.0176) [1.0736]	1.1818 (-0.0060) [0.6336]	1.1746 (-0.0052) [0.5802]	1.1807 (-0.0057) [0.7192]
	20	1.1684 (-0.0009) [0.6201]	1.3973 (0.0049) [1.0416]	1.1861 (-0.0024) [0.5941]	1.1808 (-0.0020) [0.5488]	1.1862 (-0.0022) [0.6838]
	10	1.1434 (0.0127) [0.5850]	1.3686 (0.0263) [0.9784]	1.1552 (0.0114) [0.5811]	1.1528 (0.0123) [0.5151]	1.1557 (0.0129) [0.6471]

Table S2. Simulation Results of the Static Factor Model with **J = 10** Donors.

T_0	T_1	**FACTOR**	**SC**	**REGOLS**	**NET**	**OLS**
		RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]
20	30	1.1613 (-0.0051) [0.7638]	1.2867 (-0.0034) [1.0520]	1.2304 (-0.0040) [0.7474]	1.2739 (-0.0057) [0.7799]	1.6198 (-0.0090) [2.1449]
	20	1.1555 (-0.0059) [0.7316]	1.2840 (-0.0267) [1.0055]	1.2250 (-0.0059) [0.7538]	1.2595 (0.0061) [0.7655]	1.6167 (-0.0024) [2.1752]
	10	1.1312 (-0.0091) [0.6970]	1.2532 (-0.0102) [0.9745]	1.1987 (-0.0140) [0.7263]	1.2395 (-0.0149) [0.7504]	1.5811 (-0.0052) [2.0458]
50	30	1.1043 (0.0071) [0.8100]	1.2497 (0.0032) [1.0271]	1.1369 (0.0109) [0.7619]	1.1467 (0.0111) [0.6834]	1.2086 (0.0083) [1.1120]
	20	1.0966 (0.0079) [0.7854]	1.2308 (-0.0060) [1.0024]	1.1317 (0.0024) [0.7350]	1.1407 (0.0063) [0.6546]	1.2074 (0.0048) [1.0766]
	10	1.0912 (-0.0076) [0.7318]	1.2167 (-0.0053) [0.9551]	1.1204 (-0.0075) [0.7016]	1.1281 (-0.0058) [0.6186]	1.1895 (0.0001) [1.0395]
100	30	1.0851 (0.0065) [0.7875]	1.2316 (0.0110) [1.0178]	1.1063 (0.0067) [0.7389]	1.1112 (0.0080) [0.6770]	1.1323 (0.0081) [0.9248]
	20	1.0733 (-0.0137) [0.7698]	1.2088 (0.0019) [0.9747]	1.0917 (-0.0134) [0.7290]	1.0941 (-0.0125) [0.6631]	1.1202 (-0.0122) [0.9043]
	10	1.0661 (-0.0082) [0.7524]	1.1986 (-0.0119) [0.9553]	1.0895 (-0.0096) [0.7098]	1.0934 (-0.0129) [0.6451]	1.1178 (-0.0113) [0.8864]

Table S3. Simulation Results of the Static Factor Model with **J = 15** Donors.

T_0	T_1	**FACTOR**	**SC**	**REGOLS**	**NET**	**OLS**
		RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]
20	30	1.1355 (0.0015) [0.7829]	1.2504 (-0.0022) [1.0159]	1.2171 (0.0004) [0.7942]	1.3089 (0.0012) [1.0400]	2.4509 (0.0009) [6.0410]
	20	1.1223 (0.0082) [0.7681]	1.2368 (-0.0057) [1.0051]	1.2088 (0.0130) [0.8111]	1.3010 (0.0024) [1.1466]	2.4339 (0.0089) [6.0128]
	10	1.1199 (0.0060) [0.7029]	1.2299 (0.0152) [0.9262]	1.2142 (-0.0096) [0.7565]	1.3095 (-0.0061) [1.0792]	2.4066 (-0.0104) [5.7301]
50	30	1.0834 (0.0062) [0.8168]	1.2104 (0.0041) [0.9866]	1.1219 (0.0041) [0.7623]	1.1337 (0.0041) [0.6970]	1.2756 (0.0005) [1.3377]
	20	1.0794 (0.0047) [0.7768]	1.2016 (0.0050) [0.9709]	1.1222 (0.0043) [0.7428]	1.1324 (0.0054) [0.6599]	1.2737 (0.0025) [1.2865]
	10	1.0674 (0.0001) [0.7333]	1.1873 (-0.0077) [0.8756]	1.1010 (0.0021) [0.6879]	1.1105 (0.0026) [0.6355]	1.2532 (0.0028) [1.2290]
100	30	1.0732 (0.0036) [0.8262]	1.1780 (0.0093) [0.9729]	1.0966 (0.0043) [0.7689]	1.1032 (0.0062) [0.6935]	1.1508 (0.0059) [1.0253]
	20	1.0645 (-0.0153) [0.8304]	1.1774 (-0.0091) [0.9807]	1.0905 (-0.0185) [0.7736]	1.0963 (-0.0192) [0.7014]	1.1459 (-0.0203) [1.0300]
	10	1.0627 (0.0114) [0.7628]	1.1613 (0.0090) [0.9068]	1.0829 (0.0146) [0.7123]	1.0902 (0.0149) [0.6393]	1.1378 (0.0171) [0.9584]

Table S4. Simulation Results of the Static Factor Model with **J = 20** Donors.

T_0	T_1	**FACTOR**	**SC**	**REGOLS**	**NET**	**OLS**
		RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]
20	30	1.1278 (0.0020) [0.7964]	1.2332 (-0.0040) [1.0193]	1.2183 (-0.0039) [0.8413]	1.2859 (-0.0079) [0.9901]	NA NA NA
	20	1.1244 (-0.0082) [0.7614]	1.2312 (0.0043) [0.9685]	1.2083 (-0.0054) [0.7839]	1.2712 (0.0025) [0.9067]	NA NA NA
	10	1.1054 (0.0172) [0.7525]	1.1986 (0.0063) [0.9464]	1.1987 (0.0171) [0.8319]	1.2632 (0.0169) [0.9309]	NA NA NA
50	30	1.0703 (-0.0056) [0.8408]	1.1749 (-0.0160) [0.9813]	1.1087 (-0.0070) [0.7851]	1.1247 (-0.0088) [0.7387]	1.3707 (-0.0067) [1.6546]
	20	1.0678 (0.0207) [0.8154]	1.1720 (0.0202) [0.9649]	1.1117 (0.0149) [0.7801]	1.1237 (0.0194) [0.7150]	1.3646 (0.0054) [1.5848]
	10	1.0514 (-0.0141) [0.7774]	1.1444 (-0.0264) [0.8997]	1.0846 (-0.0120) [0.7324]	1.0976 (-0.0085) [0.6672]	1.3227 (-0.0089) [1.4695]
100	30	1.0447 (-0.0039) [0.8600]	1.1422 (0.0168) [0.9745]	1.0679 (-0.0029) [0.7904]	1.0793 (-0.0007) [0.7231]	1.1626 (-0.0016) [1.1338]
	20	1.0510 (-0.0144) [0.8576]	1.1476 (-0.0185) [0.9550]	1.0715 (-0.0139) [0.7864]	1.0815 (-0.0105) [0.7234]	1.1589 (-0.0113) [1.1391]
	10	1.0304 (0.0223) [0.8118]	1.1144 (0.0166) [0.9018]	1.0505 (0.0271) [0.7425]	1.0598 (0.0270) [0.6826]	1.1397 (0.0321) [1.0761]

Table S5. Simulation Results of the Static Factor Model with **J = 25** Donors.

T_0	T_1	**FACTOR**	**SC**	**REGOLS**	**NET**	**OLS**
		RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]
20	30	1.1093	NA	1.2010	1.2554	NA
		(0.0013)	NA	(0.0090)	(0.0116)	NA
		[0.7989]	NA	[0.8648]	[0.9421]	NA
	20	1.1073	NA	1.1989	1.2708	NA
		(0.0131)	NA	(0.0108)	(0.0093)	NA
		[0.8177]	NA	[0.8516]	[0.9724]	NA
	10	1.0971	NA	1.1928	1.2447	NA
		(-0.0046)	NA	(-0.0001)	(0.0013)	NA
		[0.7589]	NA	[0.8153]	[0.8945]	NA
50	30	1.0597	1.1582	1.0996	1.1172	1.4863
		(-0.0018)	(0.0135)	(-0.0048)	(-0.0079)	(-0.0151)
		[0.8466]	[0.9590]	[0.7879]	[0.7353]	[1.9825]
	20	1.0538	1.1525	1.0993	1.1139	1.4811
		(-0.0150)	(0.0043)	(-0.0082)	(-0.0101)	(0.0033)
		[0.8350]	[0.9332]	[0.7821]	[0.7263]	[1.9856]
	10	1.0432	1.1428	1.0831	1.0997	1.4400
		(0.0062)	(-0.0038)	(0.0062)	(0.0064)	(-0.0066)
		[0.8077]	[0.9111]	[0.7625]	[0.7214]	[1.8603]
100	30	1.0378	1.1232	1.0641	1.0741	1.1959
		(0.0000)	(0.0058)	(-0.0003)	(0.0006)	(0.0044)
		[0.8501]	[0.9454]	[0.7912]	[0.7149]	[1.2244]
	20	1.0483	1.1401	1.0747	1.0842	1.2002
		(0.0110)	(0.0130)	(0.0114)	(0.0127)	(0.0147)
		[0.8655]	[0.9457]	[0.7968]	[0.7234]	[1.2330]
	10	1.0296	1.1154	1.0551	1.0648	1.1811
		(-0.0018)	(-0.0017)	(-0.0012)	(0.0001)	(-0.0040)
		[0.8083]	[0.8879]	[0.7444]	[0.6782]	[1.1463]

Table S6. Simulation Results of the Static Factor Model with **J = 30** Donors.

T_0	T_1	**FACTOR**	**SC**	**REGOLS**	**NET**	**OLS**
		RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]	RMSE (BIAS) [VARIANCE]
20	30	1.1013	NA	1.1899	1.2523	NA
		(0.0128)	NA	(0.0149)	(0.0115)	NA
		[0.8224]	NA	[0.8865]	[0.9715]	NA
	20	1.1043	NA	1.1825	1.2442	NA
		(0.0014)	NA	(0.0031)	(0.0052)	NA
		[0.8072]	NA	[0.8431]	[0.9420]	NA
	10	1.0890	NA	1.1647	1.2337	NA
		(-0.0075)	NA	(-0.0054)	(-0.0055)	NA
		[0.7761]	NA	[0.8107]	[0.8943]	NA
50	30	1.0622	1.1477	1.1039	1.1297	1.6851
		(-0.0022)	(0.0062)	(0.0009)	(-0.0012)	(0.0059)
		[0.8658]	[0.9630]	[0.8137]	[0.7852]	[2.6638]
	20	1.0508	1.1372	1.0936	1.1190	1.6716
		(-0.0016)	(-0.0084)	(-0.0009)	(-0.0004)	(-0.0006)
		[0.8700]	[0.9523]	[0.8184]	[0.7841]	[2.5964]
	10	1.0445	1.1342	1.0880	1.1061	1.6681
		(0.0006)	(-0.0043)	(0.0052)	(0.0091)	(0.0174)
		[0.8279]	[0.9218]	[0.7784]	[0.7648]	[2.5590]
100	30	1.0420	1.1178	1.0662	1.0793	1.2394
		(0.0097)	(0.0172)	(0.0107)	(0.0094)	(0.0070)
		[0.8864]	[0.9693]	[0.8258]	[0.7565]	[1.3694]
	20	1.0435	1.1237	1.0669	1.0788	1.2351
		(0.0072)	(0.0038)	(0.0054)	(0.0057)	(0.0064)
		[0.8778]	[0.9520]	[0.8077]	[0.7417]	[1.3400]
	10	1.0159	1.0890	1.0410	1.0465	1.2026
		(0.0027)	(0.0173)	(0.0041)	(0.0020)	(0.0038)
		[0.8381]	[0.8876]	[0.7678]	[0.7076]	[1.2823]

4.2. Weakly Dynamic Data Generating Processes

4.3. Dynamic Data Generating Processes

5. Applications

We consider three leading examples:

- [Abadie and Gardeazabal, 2003]
- [Abadie et al., 2010]
- [Abadie et al., 2015]

6. Conclusion

- Some concluding remark and an outlook
- Keep short, around 1-2 pages
- Natural extension: case with explanatory variables

References

[Abadie, 2021] Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425.

[Abadie et al., 2010] Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105:493–505.

[Abadie et al., 2015] Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.

[Abadie and Gardeazabal, 2003] Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93:113–132.

[Abadie and Imbens, 2002] Abadie, A. and Imbens, G. (2002). Bias-corrected matching estimators for average treatment effects. *Journal of Business and Economic Statistics*, 29.

[Abadie and Imbens, 2006] Abadie, A. and Imbens, G. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74:235–267.

[Abadie and L'Hour, 2021] Abadie, A. and L'Hour, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, 116:1–34.

[Amjad et al., 2018] Amjad, M., Shah, D., and Shen, D. (2018). Robust synthetic control. *Journal of the American Statistical Association*, pages 1–51.

[Andrews, 2003] Andrews, D. W. K. (2003). End-of-Sample Instability Tests. *Econometrica*, 71(6):1661–1694.

[Arkhangelsky et al., 2021] Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118.

[Athey et al., 2017] Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2017). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116.

[Athey and Imbens, 2016] Athey, S. and Imbens, G. (2016). The state of applied econometrics - causality and policy evaluation. *Journal of Economic Perspectives*, 31.

[Ben-Michael et al., 2021] Ben-Michael, E., Feller, A., and Rothstein, J. (2021). The augmented synthetic control method. *SSRN Electronic Journal*.

[Ben-Michael et al., 2021] Ben-Michael, E., Feller, A., and Rothstein, J. (2021). Synthetic controls with staggered adoption. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84.

[Born et al., 2019] Born, B., Müller, G. J., Schularick, M., and Sedláček, P. (2019). The Costs of Economic Nationalism: Evidence from the Brexit Experiment*. *The Economic Journal*, 129(623):2722–2744.

[Breitung and Knüppel, 2021] Breitung, J. and Knüppel, M. (2021). How far can we forecast? Statistical tests of the predictive content. *Journal of Applied Econometrics*, 36(4):369–392.

[Brodersen et al., 2015] Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. (2015). Inferring causal impact using bayesian structural time-series models. *The Annals of Applied Statistics*, 9(1):247–274.

[Cattaneo et al., 2021] Cattaneo, M., Feng, Y., and Titiunik, R. (2021). Prediction intervals for synthetic control methods*. *Journal of the American Statistical Association*, 116:1–44.

[Chernozhukov et al., 2019] Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2019). Inference on average treatment effects in aggregate panel data settings. CeMMAP working papers CWP32/19, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

[Chernozhukov et al., 2021] Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2021). An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls.

University of California at San Diego, Economics Working Paper Series qt90m9d66s, Department of Economics, UC San Diego.

[Cho, 2020] Cho, S.-W. S. (2020). Quantifying the impact of nonpharmaceutical interventions during the COVID-19 outbreak: The case of Sweden. *The Econometrics Journal*, 23(3):323–344.

[Cunningham, 2021] Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.

[Doudchenko and Imbens, 2016] Doudchenko, N. and Imbens, G. W. (2016). Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis. NBER Working Papers 22791, National Bureau of Economic Research, Inc.

[Ferman, 2021] Ferman, B. (2021). On the Properties of the Synthetic Control Estimator with Many Periods and Many Controls. *Journal of the American Statistical Association*, 116(536):1764–1772.

[Firpo and Possebom, 2018] Firpo, S. and Possebom, V. (2018). Synthetic control method: Inference, sensitivity analysis and confidence sets. *Journal of Causal Inference*, 6(2).

[Fisher, 1935] Fisher, R. A. (1971 [1935]). The design of experiments (9th ed.). *Macmillian*.

[Frangakis and Rubin, 2002] Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58(1):21–29.

[Funke et al., 2020] Funke, M., Schularick, M., and Trebesch, C. (2020). Populist leaders and the economy. ECONtribute Discussion Papers Series 036, University of Bonn and University of Cologne, Germany.

[Hahn and Shi, 2017] Hahn, J. and Shi, R. (2017). Synthetic control and inference. *Econometrics*, 5(4).

[Hainmueller et al., 2011] Hainmueller, J., Diamond, A., and Abadie, A. (2011). Synth: An r package for synthetic control methods in comparative case studies. *Journal of Statistical Software*, 42.

[Hartford et al., 2017] Hartford, J., Lewis, G., Leyton-Brown, K., and Taddy, M. (2017). Deep iv: A flexible approach for counterfactual prediction. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 1414–1423. JMLR.org.

[Harvey and Thiele, 2020] Harvey, A. and Thiele, S. (2020). Cointegration and control: Assessing the impact of events using time series data. *Journal of Applied Econometrics*, 36.

[Kellogg et al., 2021] Kellogg, M., Mogstad, M., Pouliot, G. A., and Torgovitsky, A. (2021). Combining matching and synthetic control to tradeoff biases from extrapolation and interpolation. *Journal of the American Statistical Association*, 116(536):1804–1816. PMID: 35706442.

[Kuosmanen et al., 2021] Kuosmanen, T., Zhou, X., Eskelinen, J., and Malo, P. (2021). Design Flaw of the Synthetic Control Method. MPRA Paper 106328, University Library of Munich, Germany.

[Martin et al., 2012] Martin, V., Hurn, S., and Harris, D. (2012). *Econometric Modelling with Time Series: Specification, Estimation and Testing*. Themes in Modern Econometrics. Cambridge University Press.

[Muhlbach and Nielsen, 2019] Muhlbach, N. S. and Nielsen, M. S. (2019). Tree-based Synthetic Control Methods: Consequences of moving the US Embassy. Papers 1909.03968, arXiv.org.

[Neyman, 1923] Neyman, J. (1923). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science 5*, 4:465–472.

[Rosenbaum and Rubin, 1983] Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.

[Rubin, 1974] Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

[von Brzeski et al., 2015] von Brzeski, V., Taddy, M., and Draper, D. (2015). Causal inference in repeated observational studies: A case study of ebay product releases.

Appendix

- Some proofs would be nice.